

University of Groningen

Some estimates of the optimum inductive method

Kuipers, Theo A.F.

Published in:
Erkenntnis, vol. 24, 37-46

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
1986

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Kuipers, T. A. F. (1986). Some estimates of the optimum inductive method. *Erkenntnis*, vol. 24, 37-46, 24, 37-46.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

THEO A. F. KUIPERS

SOME ESTIMATES OF THE OPTIMUM INDUCTIVE METHOD

1. INTRODUCTION

It is well known that Carnap's *Continuum of Inductive Methods* (1952) contains the calculation of an optimum value of the crucial parameter λ of the continuum. However, this value is based on the fundamental unknowns in a proper inductive situation, viz., the underlying objective probabilities. Hence, this optimum value is only a theoretical construct, without practical use-value.

The leading idea of this paper is that it may nevertheless be possible to estimate the optimum value on the basis of the available evidence.

Although this idea is very plausible, it must be confessed that our results are not far-reaching. In particular, although we can formulate some appealing estimates, we did not succeed in evaluating the quality of these estimates as estimates. Hopefully, statisticians can come further with it. However, we do also not exclude, to say the least, that the present line of research is a dead alley. But, if so, somebody had to walk in it.

In Section 2 we will summarize, in our terms, Carnap's derivation of the existence of an optimum value of λ , always indicated by λ^Δ . In Section 3 we will define a limit-process in order to approach λ^Δ , and we will investigate the main aspects of the limit-behaviour. However, all other questions of evaluation turn out to be too difficult. In Section 4 we will come somewhat further in dealing with the simplest 'one-step estimate' and some of its variants. We conclude, in Section 5, with some final remarks. The Appendix contains a table of relevant expectations and variances, as far as we succeeded in calculating them.

2. THE OPTIMUM INDUCTIVE METHOD

In accordance with the pleas in our (1978) and (1984a) we will use the logico-mathematical way of presentation, instead of the logico-linguistic way.

In the present context the underlying probability process is supposed

to be straightforward multinomial. Let $K = \{Q_1, \dots, Q_i, \dots, Q_k\}$ be the set of (elementary) outcomes ($2 \leq k < \infty$), with objective probability r_i ($r_i \geq 0, \sum r_i = 1$). For a sequence of n outcomes $e_n \in K^n$, the number of occurrences of Q_i in e_n is indicated by $\underline{n}_i(e_n)$, or simply \underline{n}_i if e_n is fixed in the context, which will be the rule. All random variables depending on a fixed e_n will be underlined. Summations will be over $i = 1, \dots, k$, if not indicated otherwise.

An inductive method belonging to Carnap's continuum, or simply a λ -method, is a probability system, based on a fixed real number λ , $0 \leq \lambda \leq \infty$, with special values

$$(1) \quad p_i =_{\text{df}} p(Q_i/e_n) = (\underline{n}_i + \lambda/k)/(n + \lambda)$$

The extreme method with $\lambda = 0$ is called *the straight rule*, the extreme method with $\lambda = \infty$ is called *the noninductive method*.

The existence of an optimum method is not difficult to prove. Let us call

$$(2) \quad \underline{d}_i =_{\text{df}} p_i - r_i$$

the error of p_i with respect to r_i . As is well known, and easy to prove, the expected square error $E(\underline{d}_i^2)$ is equal to the sum of the square of the expected error ($E^2(\underline{d}_i)$) and the variance of the error ($\sigma^2(\underline{d}_i)$). With reference to the Appendix, it is now easy to calculate the sum-total of the expected square errors:

$$(3) \quad S(\lambda, n) =_{\text{df}} \sum E(\underline{d}_i^2) = [\lambda^2(R^\Delta - 1/k) + n(1 - R^\Delta)]/(\lambda + n)^2$$

with

$$(4) \quad R^\Delta =_{\text{df}} \sum r_i^2 \quad 1/k \leq R^\Delta \leq 1$$

This R^Δ is an index of order for which it is easy to prove that it reaches its minimum $1/k$ if all r_i are equal, i.e., $1/k$, and its maximum 1 if one r_i is 1 (and all others are 0).

$S(\lambda, n)$ is of course the plausible measure of the (global) quality of a particular λ -method, at least from the omniscient point of view. By standard procedures it is easy to prove the following crucial proposition.

THEOREM. $S(\lambda, n)$ is minimal for

$$(5) \quad \lambda^\Delta =_{\text{df}} (1 - R^\Delta)/(R^\Delta - 1/k)$$

It is important to note that λ^Δ is independent of n . For this reason it

makes sense to call λ^Δ the optimum value of λ and the λ^Δ -method the optimum (inductive) method.

The way in which λ^Δ depends on R^Δ and k is as follows

- (6) if $r_i = 1/k$ for all i , then $R^\Delta = 1/k$ and $\lambda^\Delta = \infty$,
 if $r_i = 1$ for some(one) i , then $1 = R^\Delta$ and $0 = \lambda^\Delta$,
 otherwise $1 > R^\Delta > 1/k$ and $0 < \lambda^\Delta < \infty$.

Hence, we see that the optimum method is an extreme one if and only if the index of order is extreme.

Although R^Δ may play a role as index of order or index of homogeneity,¹ the corresponding optimum λ^Δ cannot play a role in a proper inductive context. For in such a context, the r_i 's and hence R^Δ and λ^Δ are unknown. This is true, even if there is some evidence e_n available. However, in this case we can try to estimate λ^Δ on the basis of e_n . For the special case that e_n is *uniform*, i.e., all \underline{n}_i are equal to n/k (including 'zero evidence'), all λ -methods lead to the same special values, viz., $1/k$, and hence estimation of λ^Δ is then not interesting. Therefore we will generally assume *nonuniform* evidence.

3. ESTIMATION BY A LIMIT-PROCESS

The idea behind our first proposal is a simple one. Assume nonuniform e_n . Starting from an arbitrary (finite) value for λ , we can calculate the corresponding p_i 's and then 'the corresponding values' of R^Δ and λ^Δ , respectively. After such a cycle we can start over again, and so infinitely many times. In a scheme we get:

Limit-process

Start	λ_0 arbitrary	$0 \leq \lambda_0 < \infty$
	$\downarrow_{m=1}$	
Restart: $m = 2, 3 \dots$	$p_i(m) = \frac{\underline{n}_i + \underline{\lambda}_{m-1}/k}{n + \underline{\lambda}_{m-1}}$	$0 \leq p_i(m) \leq 1$
	\downarrow	
	$\underline{R}_m = \sum p_i^2(m)$	$1 \geq \underline{R}_m > 1/k$
	\downarrow	
	$\underline{\lambda}_m = \frac{1 - \underline{R}_m}{\underline{R}_m - 1/k}$	$0 \leq \underline{\lambda}_m < \infty$

Formulating this limit-process is one thing,² evaluating it another. The main questions of evaluation are obviously:

- what is the limit-behaviour of this process?
- what is the quality of the limit of $\underline{\lambda}_m$ as estimate of λ^Δ ?

Fortunately we can give now a global answer to the first question. The difficulty of answering the second question will force us to less complicated proposals.

We will use the following abbreviations:

$$(7) \quad \begin{aligned} \underline{R}^0 &= \underline{R}_{m=1}^{\lambda=0} = \Sigma(\underline{n}_i/n)^2 & 1 \geq \underline{R}^0 > 1/k \\ \underline{\lambda}^0 &= \underline{\lambda}_{m=1}^{\lambda=0} = (1 - \underline{R}^0)/(\underline{R}^0 - 1/k) & 0 \leq \underline{\lambda}^0 < \infty \\ \underline{C}^0 &= \frac{k-1}{n^2(k\underline{R}^0-1)} & \frac{1}{n^2} \leq \underline{C}^0 < \infty \end{aligned}$$

It is not difficult to derive that $\underline{\lambda}_{m+1}$ is the following function of $\underline{\lambda}_m$.

$$(8) \quad \underline{\lambda}_{m+1} =_{\text{df}} f(\underline{\lambda}_m) = \underline{C}^0(\underline{\lambda}_m + n)^2 - 1$$

Some features of the parabole $f(\lambda)$ are directly clear. It reaches its minimum -1 for $\lambda = -n$. Except for extreme nonuniform evidence, $f(\lambda) = 0$ for two negative values of λ , i.e., $-n \pm \sqrt{1/\underline{C}^0} < 0$, and $f(\lambda = 0) = \underline{\lambda}^0 > 0$. For extreme nonuniform evidence, i.e., $\underline{n}_i = n$ for some i , and hence $\underline{R}^0 = 1$, we have $f(\lambda = 0) = 0 = f(\lambda = -2n)$.

The limit-behaviour is now of course governed by the quadratic equation

$$(9) \quad f(\lambda) = \underline{C}^0(\lambda + n)^2 - 1 = \lambda$$

or, equivalently,

$$\underline{C}^0\lambda^2 + (2n\underline{C}^0 - 1)\lambda + n^2\underline{C}^0 - 1 = 0$$

It is easy to check that 0 is a root if and only if we have extreme nonuniform evidence and that, if there are two non-zero roots, they are both positive or both negative. The result of the limit-process can now be described in four possible cases.

Case 1. Two non-negative roots λ_{left} and λ_{right} , $0 \leq \lambda_{\text{left}} < \lambda_{\text{right}}$

$$\begin{array}{ll} \text{if } 0 \leq \lambda_0 < \lambda_{\text{right}} & \text{then } \underline{\lambda}_m \rightarrow \lambda_{\text{left}} \\ \text{if } \lambda_0 = \lambda_{\text{right}} & \text{then } \underline{\lambda}_m = \lambda_{\text{right}} \\ \text{if } \lambda_{\text{right}} < \lambda_0 & \text{then } \underline{\lambda}_m \rightarrow \infty \end{array}$$

Case 2. One unique non-negative root λ_{uni} , $0 \leq \lambda_{\text{uni}}$

if $0 \leq \lambda_0 \leq \lambda_{\text{uni}}$ then $\underline{\lambda}_m \rightarrow \lambda_{\text{uni}}$
 if $\lambda_{\text{uni}} < \lambda_0$ then $\underline{\lambda}_m \rightarrow \infty$

Case 3. One negative root and root zero

if $\lambda_0 = 0$ then $\underline{\lambda}_m \rightarrow 0$
 if $\lambda_0 > 0$ then $\underline{\lambda}_m \rightarrow \infty$

Case 4. All other cases

(no roots, two negative roots, or a unique negative root)

for all λ_0 , $0 \leq \lambda_0 < \infty$ $\underline{\lambda}_m \rightarrow \infty$

It should however be noted that negative roots (Case 3 and subcases of Case 4) are exceptional; moreover they are always between 0 and -1 . For non-extreme nonuniform evidence the main cases are two positive roots (Case 1), via a unique positive root (Case 2), to no roots at all (Case 4). The graph shows the limit-behaviour in Case 1 for the three interesting regions for the initial value λ_0 . Note that λ_{left} is an attractor and that λ_{right} is an unstable fixed-point. All other cases can be demonstrated in the same way.

So far for the answer to the (first) question about the limit-behaviour. As to the (second) question about the quality of the limit of $\underline{\lambda}_m$ as estimate of λ^A , we can distinguish at least three subquestions, which will be formulated for Case 1, assuming in addition that $0 \leq \lambda_0 < \lambda_{\text{right}}$, and hence that $\underline{\lambda}_m \rightarrow \lambda_{\text{left}}$:

- (i) is λ_{left} an unbiased estimate of λ^A , i.e., $E(\lambda_{\text{left}}) = \lambda^A$?
- (ii) if biased, does $E(\lambda_{\text{left}})$ go to λ^A for increasing n ?
- (iii) what is the (behaviour of the) variance of λ_{left} ?

With respect to the first question, it should be remarked that it appears very difficult to calculate the expectation value of λ_{left} . However, using probability inequalities, one can establish that the estimate is generally not unbiased. We do not know the answer on the second and third question.

In view of the difficulty of these quality-questions we decided to look for less complicated estimates, viz., at the beginning of the limit-process, leaving the process itself for what it is.

This graph is based on $k=5$,
 $n=9=6+3+0+0+0$, leading to the function

$$f(\lambda) = \left(\frac{\lambda+9}{6}\right)^2 - 1$$

and the quadratic equation
 $f(\lambda) = \lambda$:

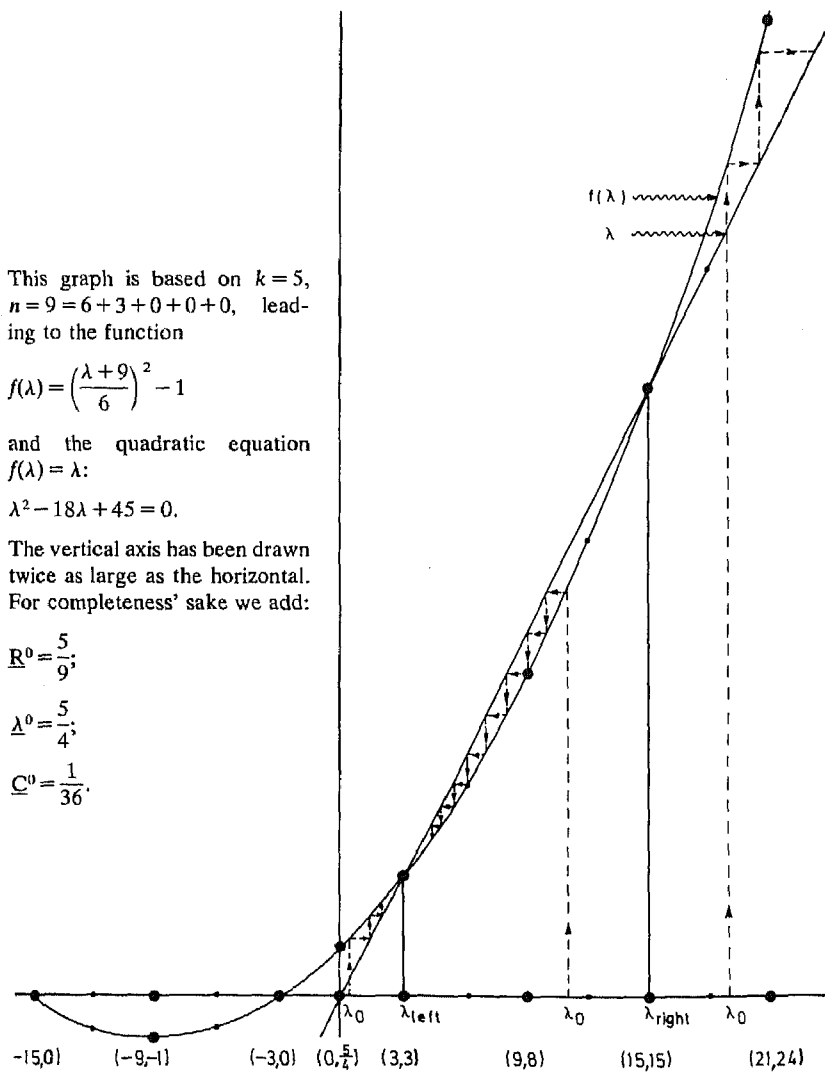
$$\lambda^2 - 18\lambda + 45 = 0.$$

The vertical axis has been drawn
 twice as large as the horizontal.
 For completeness' sake we add:

$$R^0 = \frac{5}{9};$$

$$\underline{\lambda}^0 = \frac{5}{4};$$

$$\underline{C}^0 = \frac{1}{36}.$$



4. ONE-STEP ESTIMATES

We start with repeating the definitions of \underline{R}^0 and $\underline{\lambda}^0$ in (7):

$$\begin{aligned}\underline{R}^0 &= \Sigma(\underline{n}_i / \underline{n}_i)^2, \quad 1 \geq \underline{R}^0 > 1/k \\ \underline{\lambda}^0 &= (1 - \underline{R}^0) / (\underline{R}^0 - 1/k), \quad 0 \leq \underline{\lambda}^0 < \infty\end{aligned}$$

As is easy to verify, $\underline{\lambda}^0$ is the result of applying the limit-process only for one step, starting with $\lambda_0 = 0$. What about $\underline{\lambda}^0$ as estimate of λ^Δ ?

Let us first look at \underline{R}^0 as estimate of R^Δ . From the Appendix we see that it is a biased estimate, which approaches R^Δ however, with increasing n . Its variance, which is complicated to calculate, does not give very much insight.

The expectation and variance of $\underline{\lambda}^0$ were too difficult to calculate. Now there is however a plausible quality question of comparative nature: is the ' $\underline{\lambda}^0$ -(probability-)system', on the average, superior to the straight rule ($\lambda = 0$)? Remember Carnap's result (3) in Section 2 that the sum-total $S(\lambda, n)$ of the expected square errors is a function of λ , n and of course R^Δ and k . Writing $\lambda(R) =_{\text{df}} (1 - R(\lambda)) / ((R(\lambda) - 1/k))$, we will indicate $S(\lambda, n)$ in terms of R , i.e., $R(\lambda)$, by $T(R, n)$. Hence, we have then e.g., $T(\underline{R}^0, n) = S(\underline{\lambda}^0, n)$ and $T(1, n) = S(0, n) = (1 - R^\Delta)/n$. What we would like to compare is $T(\underline{R}^0, n)$ and $T(1, n)$.

In particular, the interesting question is:

$$\underline{A} =_{\text{df}} A(\underline{R}^0) = T(1, n) - T(\underline{R}^0, n) \geq 0?$$

After some manipulation, it is possible to divide the inequality by the factor $(1 - \underline{R}^0)$, leading to a question of the form:

$$\underline{B} =_{\text{df}} B(\underline{R}^0) = a\underline{R}^0 + b \geq 0?$$

where a and b depend only on R^Δ , k and n .

Although it is easy to check that this inequality will have exceptions, it is rather difficult to give some insight in these exceptions. However, it is not too difficult to prove that the expectation $E(\underline{B}) = aE(\underline{R}^0) + b$ is straightforward non-negative. Unfortunately, due to the division by $(1 - \underline{R}^0)$, we may not conclude that the expectation $E(\underline{A})$ is always non-negative.

This suggests the question whether there is perhaps even some positive integer j such that $E(\underline{B}) - j\sigma(\underline{B}) \geq 0$, or equivalently,

$$aE(\underline{R}^0) + b - j|a|\sigma(\underline{R}^0) \geq 0?$$

For $j=2$ we found only that the resulting inequality is not without exceptions.

Thanks to some suggestions of Roberto Festa³ we can add some variants of the foregoing one-step estimate of R^Δ . The first one is an unbiased modification of \underline{R}^0 :

$$\hat{\underline{R}} =_{\text{df}} (n\underline{R}^0 - 1)/(n - 1).$$

That $\hat{\underline{R}}$ is an unbiased estimate of R^Δ , which is easy to prove, makes it of course preferable in this respect to \underline{R}^0 . However, everything we have said about the comparison of the corresponding $\underline{\lambda}^0$ -system ($\underline{\lambda}^0 = \lambda(\underline{R}^0)$) with the straight rule, can be proved also about the comparison of the $\hat{\underline{\lambda}}$ -system ($\hat{\underline{\lambda}} = \lambda(\hat{\underline{R}})$) with that rule.

Another variant of \underline{R}^0 can be obtained as follows. \underline{R}^0 was a one-step estimate starting from $\lambda_0 = 0$. If we start with $\lambda_0 = k$ we get

$$\underline{R}^* =_{\text{df}} \Sigma(\underline{n}_i + 1)^2 / (n + k)^2$$

Again we can also make a modification of \underline{R}^* in a similar way as $\hat{\underline{R}}$ is a modification of \underline{R}^0 , viz.,

$$\hat{\underline{R}} =_{\text{df}} (n\underline{R}^* - 1)/(n - 1)$$

However, not only \underline{R}^* but also $\hat{\underline{R}}$ is now a biased estimate of R^Δ . Moreover, it is easy to prove that there is no linear function of \underline{R}^* that leads to an unbiased estimate.

Unfortunately, comparison of the two corresponding probability systems with the straight rule is now still more complicated than it already was in the first two cases.

5. FINAL REMARKS

As announced in the Introduction the results of this paper are not far-reaching. A question that we did not yet touch is: what kind of probability system results if we substitute for λ in (1) an estimate of λ^Δ ? From the definition of a λ -method it is already clear that the result is not a λ -method, for an estimate of λ^Δ on the basis of e_n cannot be a *fixed* real number.

It is equally easy to see that the resulting system, for $k \geq 3$, cannot have the property that $p(Q_i/e_n)$ is only a function of n and n_i , being a structural property of all λ -systems.

Whatever other properties the resulting systems may or may not

have, it is clear that in the special case of $\lambda_0 = 0$ we get limit- and one-step-systems which are essentially parameter-free, i.e., we do not have to choose a priori a value for some parameter, which is a crucial characteristic of all λ -methods, except of the extreme, straight rule and of the other extreme, viz., the non-inductive method.⁴

APPENDIX: TABLE OF EXPECTATIONS AND VARIANCES

Random variable	Expectation	Variance
\underline{x}	$E(\underline{x})$	$\sigma^2(\underline{x}) =_{\text{df}} E(\underline{x} - E(\underline{x}))^2$ $= E(\underline{x}^2) - E^2(\underline{x})$
$a\underline{x} + b$	$aE(\underline{x}) + b$	$a^2\sigma^2(\underline{x})$
\underline{n}_i	$r_i n$	$r_i(1 - r_i)n$
$\underline{p}_i = \frac{\underline{n}_i + \lambda/k}{n + \lambda}$	$\frac{r_i n + \lambda/k}{n + \lambda}$	$\frac{r_i(1 - r_i)n}{(n + \lambda)^2}$
$\underline{d}_i = \underline{p}_i - r_i$	$\frac{\lambda/k - \lambda r_i}{n + \lambda}$	$\sigma^2(\underline{p}_i)$
$\underline{R}^0 = \Sigma \left(\frac{\underline{n}_i}{n} \right)^2$	$\frac{(n-1)R^\Delta + 1}{n}$ ^a	$\frac{2(n-1)}{n^3} [2(n-2)\Sigma r_i^3 +$ $R^\Delta - (2n-3)(R^\Delta)^2]$ ^b
$\hat{\underline{R}} = \frac{n\underline{R}^0 - 1}{n-1}$	R^Δ	$\left(\frac{n}{n-1} \right)^2 \sigma^2(\underline{R}^0)$
$\underline{R}^* = \Sigma \left(\frac{\underline{n}_i + 1}{n+k} \right)^2$ $= \frac{n^2 \underline{R}^0 + 2n + k}{(n+k)^2}$	$\frac{n(n-1)R^\Delta + 3n + k}{(n+k)^2}$	$\left(\frac{n}{n+k} \right)^4 \sigma^2(\underline{R}^0)$
$\hat{\underline{R}} = \frac{n\underline{R}^* - 1}{n-1}$	$\frac{n^2(n-1)R^\Delta + 2n^2 - nk - k^2}{(n-1)(n+k)^2}$	$\frac{n^6}{(n-1)^2(n+k)^4} \sigma^2(\underline{R}^0)$

^a Recall that R^Δ is defined as Σr_i^2 .

^b Calculated by using the method of moment generating functions (see e.g., Johnson and Kotz, *Distributions in Statistics*, Wiley, 1969).

NOTES

¹ That R^Δ is an index of order or homogeneity (and, hence, $(1 - R^\Delta)$ an index of disorder or heterogeneity) was already argued by C. Gini in 1912, but apparently this was unknown to Carnap. This fact and many other interesting information about R^Δ and λ^Δ

has been presented by Roberto Festa (1984). E.g., about the foundation of $(1 - R^\Delta)$ as an index of heterogeneity, about the factual use of R^Δ in different sciences, about the relation between R^Δ and λ^Δ , and about normalized versions of both. Finally, Festa shows that Carnap's derivation of λ^Δ can be reinterpreted as the derivation of the optimum way of approaching the truth.

² I formulated this limit-process, with $\lambda_0 = 0$, already in my doctoral thesis of 1971 and a note of 1972. In a discussion in 1982, Jaakko Hintikka turned out to be interested in the idea, because the limit-process with $\lambda_0 = 0$ leads to an inductive system which is free of a priori parameters. The note appeared recently in its original form (Kuipers, 1984b). This course of events stimulated also some new research, leading to the present paper.

³ In particular, Festa explained in a letter to me 'after Siena' that close inspection of Good (1953), partly undertaken by Herdan (1958), shows that Good's article includes some estimates of R^Δ . In the present paper, \hat{R} corresponds to Good's $\hat{c}_{2,0}$ (see his Section 6) and \hat{R} is one possible specification of his $\hat{c}_{2,0}$ (see his Sections 3 and 6).

⁴ I am grateful to Otto Kardaun for his help in solving some specific mathematical problems.

REFERENCES

- Carnap, R.: 1952, *The Continuum of Inductive Methods*, University of Chicago Press, Chicago.
- Festa, R.: 1984, 'Inferenza Predittiva e Misure di Eterogeneità', *Statistica*, **XLIV**, 111-129.
- Good, I. J.: 1953, 'The Population Frequencies of Species and the Estimation of Population Parameters', *Biometrika* **40**, 237-264.
- Herdan, G.: 1958, 'The Mathematical Relation between Greenberg's Index of Linguistic Diversity and Yule's Characteristic', *Biometrika* **45**, 268-270.
- Kuipers, T.: 1978, *Studies in Inductive Probability and Rational Expectation*, *Synthese Library* 123, Reidel, Dordrecht.
- Kuipers, T.: 1984a, 'What Remains of Carnap's Program Today?', *Probability, Statistics and Inductive Logic*, Proc. Int. Conf. Luino, 1981, special issue of *Epistemologia*, eds. E. Agazzi, D. Costantini, pp. 121-152.
- Kuipers, T.: 1984b, 'An Approximation of Carnap's Optimum Estimation Method', *Synthese* **61**, 361-362.

Manuscript received 8 May 1985

Rijksuniversiteit Groningen
 Filosofisch Instituut
 Westersingel 19
 9718 CA Groningen
 The Netherlands